SemeAd 2023
XXVI Seminários em Administração

07, 08, 09 e 10 de novembro de 2023
ISSN 2177-3866

EXPLORING THE TEXTUAL INFORMATION OF PATENT DOCUMENTS USING ARTIFICIAL INTELLIGENCE

KÁTIA CINARA TREGNAGO CUNHA
UNIVERSIDADE NOVE DE JULHO (UNINOVE)

CRISTINA DAI PRÁ MARTENS
UNIVERSIDADE NOVE DE JULHO (UNINOVE)

CARLA BONATO MARCOLIN
UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)

Agradecimento à orgão de fomento:
The authors would like to thank the Research Support Fund - FAP UNINOVE.

SemeAd 2023
XXVI Seminários em Administração

07, 08, 09 e 10 de novembro de 2023
ISSN 2177-3866

EXPLORING THE TEXTUAL INFORMATION OF PATENT DOCUMENTS USING ARTIFICIAL INTELLIGENCE

Introdução

Collections of patent documents are an important source of knowledge for research and innovation and for the productive sector. To analyze large amounts of unstructured data, artificial intelligence has allowed to improve the extraction of information and transform it into an understandable form. Patent documents differ from general documents because they present a specific form of expression, providing structured and unstructured fields, in the form of texts and/or images (Sun et al., 2021).

Problema de Pesquisa e Objetivo

The article seeks to answer the following research question: How do artificial intelligence (AI) tools extract knowledge from textual fields of patent documents? The objective of the study is to analyze the accumulated knowledge about AI techniques for mining textual fields of patent documents, identifying the strategies, methods, approaches and patent fields explored.

Fundamentação Teórica

In text mining, keyword extraction is critical, having a significant influence on subsequent analytical results (Hu et al., 2018). The extraction of key terms from unstructured fields of patent documents is a challenging task, due to the form of expression in different domains of technology (Liwei, 2022). In massive textual data analysis through AI, unstructured text content is converted into a formal, machine-readable representation for automatic processing, improving information extraction.

Discussão

The Systematic Literature Review (SLR) selected 69 articles, published from 2018 onwards, by consulting the Scopus and Web of Science databases. The keyword extraction method (supervised, unsupervised or semi-supervised); the text mining approach (machine learning, deep learning, Natural Language Processing (NLP)); the type of text mining approach (keywords, Subject-Action-Object, Property-Function), the methodology and the patent field explored.

Conclusão

In the field of AI, NLP is still predominantly used to extract patent information. However, there is a growth in the use of deep learning and machine learning techniques, mainly due to the difficulties in processing the large amount of data. About the keyword extraction method, unsupervised learning stands out. Mining tools explore, with greater emphasis, one to three fields of patent textual data selected from title, abstract, claims, detailed description or technical field.

Referências Bibliográficas

Hu, J., Li, S., Yao, Y., Yu, L., Yang, G., & Hu, J. (2018). Patent Keyword Extraction Algorithm Based on Distributed Representation for Patent Classification. Entropy, 20(2). Sun, Y.-D., Cao, G.-Z., Gao, C., Yang, W.-D., Han, W.-P., & Wang, K. (2021). Extraction and Modeling of Chinese Patent Information for Technical Advancement Evaluation. IFIP Advances in Information and Communication Technology, 635 IFIP, 127–140. Xie, Z., & Miyazaki, K. (2013). Evaluating the effectiveness of keyword search strategy for patent identification. World Patent Information, 35(1), 20–30.